

Exhibit A

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,

Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)

TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

Plaintiffs,

v.

NANDITA BERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

United States' Database Matching Protocol

The United States hereby sets out its protocol for comparisons among and between the Texas databases and the federal databases. The matching process proceeds in four parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, one-to-many matches are conducted between databases. *Fourth*, Texas data concerning deceased individuals is appended to matching output. Each of the four parts is divided into multiple stages, which in turn are divided into concrete steps.

The database preparation and standardization steps below apply to both the matches requested by the United States and all Plaintiffs, as well as the matches requested by Defendants.

GLOSSARY

The following definitions apply to the terms used in this document.

- **Disability database:** Any federal database containing records that indicate whether an individual has a disability that would permit that individual apply for an exemption from the SB 14 identification requirement.
- **Identification database:** Any Texas or federal database containing records that indicate whether an individual has identification required by SB 14 in most cases to cast an in-person ballot
- **Protocol:** The instructions to prepare data for matching and to conduct multiple matching sweeps between databases according to specified search criteria.
- **Sweep:** A comparison of a set of database fields in the TEAM database against a set of database fields in a disability database or an identification database.
- **TEAM database:** The Texas Election Administration Management database, the state database that contains records of registered voters.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from TEAM, Identification & Disability Databases

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).
- Step 1.1.7:** Extract unique record identifier (such as VUID in the TEAM database).

Stage 2: Separate Valid Identification and Disability Records

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).
- Step 1.2.2:** Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.
- Step 1.2.3:** Separate the identification databases produced by the State of Texas into separate files for each form of state identification at issue, namely driver licenses, identification cards, licenses to carry concealed handguns, and election identification certificates.

Stage 3: Diagnostics

- Step 1.3.1:** Report the frequency of missing values for each field.
- Step 1.3.2:** Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.
- Step 1.3.3:** Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

- Step 1.4.1:** Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.
- Step 1.4.2:** For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.
- Step 1.4.3:** Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.
- Step 1.4.4:** Code all missing values as blank fields.

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.¹

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

¹ The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Stage 8: Standardize Address

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.²
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

² For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

Stage 10: Identical Records

Step 1.10.1: For the TEAM database, for the small number of records with different VUID but identical first and last name, gender, residential address number and ZIP, date of birth, and SSN, treat these records as identical.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.1.1:** Create Combination A: First Name + Last Name + Gender + Date of Birth + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.³
- Step 2.1.2:** Create Combination B: Last Name + Gender + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.3:** Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.4:** Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.5:** Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.
- Step 2.1.6:** Create Combination F: First Name + Last Name + Gender + Date of Birth.
- Step 2.1.7:** Create Combination M: Texas Driver License Number (where available).

³ For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.⁴
- Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.
- Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.
- Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.
- Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.
- Step 2.2.6:** Full Social Security Number.

Stage 3: Establish Identifier Uniqueness

- Step 2.3.1:** Within the TEAM database, determine the uniqueness of each primary and secondary identifier variable and mark accordingly. *E.g.*, if only one record has the string <JEANSMITH0101195001237777> for Combination A, mark that record as unique for Combination A. By contrast, if multiple records have the string <JOHNSMITHA0101950> for Combination G, mark all such records as non-unique for Combination G.
- Step 2.3.2:** Within the identification and disability databases, generate a field that establishes the uniqueness of each identifier variable. For federal databases, for each combination, generate a field that establishes

⁴ Only for the State Department, create three further variations of Combination G using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH0101195001237777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identification or disability database. For federal databases, use only the subset of records with Texas addresses in the identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identification or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identification or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identification databases produced by the State of Texas.

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identification or disability database. For federal databases, use only the subset of records with Texas addresses in the identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identification or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number⁵ in the TEAM database against the equivalent combination/field in the identification or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identification or disability database.

⁵ The full social security number is not created as a separate “combination” as it is its own field stored within the TEAM database under the field name “ssn”.

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identification or disability database.

Stage 3: United States' Nationwide Federal Database Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identification or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identification or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identification or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination

K against Combination G, Combination K, and Combination L in the nationwide identification or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identification or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identification or disability database.⁶

⁶ Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

PART IV: DATA GATHERING

Stage 1: Gather Information Regarding Deceased Individuals

Step 4.1.1: In all instances in which a unique match was achieved between the TEAM database and the Texas Driver License, Texas Identification Card, and Texas Concealed Handgun database, determine whether the identification record at issue has been flagged as deceased in the Texas identification database.

Step 4.1.2: Where the driver license, concealed handgun license, or Texas identification card record at issue has been flagged as deceased, append the “deceased” flag to the TEAM record to which the unique match has been made.